

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/12, C07K 14/47, A61K 38/17,</b> <b>C12N 15/10, C12Q 1/68</b>	<b>A2</b>	<b>(11) International Publication Number: WO 98/45436</b> <b>(43) International Publication Date: 15 October 1998 (15.10.98)</b>
<b>(21) International Application Number:</b> PCT/US98/06955 <b>(22) International Filing Date:</b> 10 April 1998 (10.04.98)  <b>(30) Priority Data:</b> 08/838,821 10 April 1997 (10.04.97) US  <b>(71) Applicant:</b> GENETICS INSTITUTE, INC. [US/US]; 87 CambridgePark Drive, Cambridge, MA 02140 (US).  <b>(72) Inventors:</b> JACOBS, Kenneth; 151 Beaumont Avenue, Newton, MA 02160 (US). MCCOY, John, M.; 56 Howard Street, Reading, MA 01867 (US). LAVALLIE, Edward, R.; 113 Ann Lee Road, Harvard, MA 01451 (US). RACIE, Lisa, A.; 124 School Street, Acton, MA 01720 (US). MERBERG, David; 2 Orchard Drive, Acton, MA 01720 (US). TREACY, Maurice; 93 Walcott Road, Chestnut Hill, MA 02167 (US). SPAULDING, Vikki; 11 Meadowbank Road, Billerica, MA 01821 (US). AGOSTINO, Michael, J.; 26 Wolcott Avenue, Andover, MA 01810 (US).  <b>(74) Agent:</b> SPRUNGER, Suzanne, A.; Genetics Institute, Inc., 87 CambridgePark Drive, Cambridge, MA 02140 (US).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> SECRETED EXPRESSED SEQUENCE TAGS (sESTs)  <b>(57) Abstract</b>  Secreted expressed sequence tags (sESTs) isolated from a variety of human tissue sources are provided.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## SECRETED EXPRESSED SEQUENCE TAGS (sESTs)

5 FIELD OF THE INVENTION

The present invention provides novel polynucleotides which are expressed sequence tags (ESTs) for secreted proteins.

## BACKGROUND OF THE INVENTION

10 Gargantuan efforts have been employed by various investigational projects to randomly sequence portions of naturally-occurring cDNAs. The rationale behind this approach to identification and sequencing genes is founded in two basic principles: (1) that transcribed cDNAs represent the product of the most important genes, namely those that are actually expressed *in vivo*, and (2) that efforts to sequence genes and other portions of the genome of  
15 target organisms which are not actually expressed wastes substantial effort on areas not likely to yield genetic information of therapeutic importance. Thus, the high-throughput sequencing efforts focus on only those portions of the genome which are expressed. The randomly produced cDNA sequences represent "expressed sequence tags" or "ESTs", which identify and can be used as probes for the longer, full-length cDNA or genomic sequence from which they  
20 were transcribed.

Although this "shortcut" approach to genomic sequencing presents savings of effort compared to sequencing of the complete genome, it still produced a vast array of ESTs which may not be directly useful as protein therapeutics. To date, the majority of protein-related drug discovery has focused on the use of secreted proteins to produce a desired therapeutic effect.

25 Since the EST approach theoretically identifies all expressed proteins, it produces an EST library which contains a mixture of secreted proteins (such as hormones, cytokines and receptors) and non-secreted proteins (such as, for example, metabolic enzymes and cellular structural proteins), without identifying which ESTs correspond to proteins falling into either category. As a result, these methods are not optimally tailored to the needs of investigators

30 searching for secreted proteins because they must separate the secreted "wheat" from the non-secreted "chaff", wasting effort and resources in the process.

Co-assigned U.S. Patent No. 5,536,637, which is incorporated herein by reference, provides methods for focusing genomic sequencing efforts on sequences encoding the secreted proteins which are of most interest for identification of protein therapeutics. The '637 patent  
35 discloses a "signal sequence trap" which selectively identifies ESTs for secreted

proteins, namely "secreted expressed sequence tags" or "sESTs". It is to these sESTs that the present invention is directed.

### SUMMARY OF THE INVENTION

5       The present invention provides for sESTs isolated from a variety of human RNA/cDNA sources.

In preferred embodiments, the present invention provides an isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

10       SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ  
ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID  
NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ  
ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ  
ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ  
ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ  
15       ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ  
ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ  
ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ  
ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ  
ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ  
20       ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ  
ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ  
ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ  
ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ  
ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ  
25       ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ  
ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ  
ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ  
ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100,  
SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID  
30       NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109,  
SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID  
NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118,  
SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID  
NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127,

TGGCTCACTC	AATGACCTCC	AGTTCTTTAG	ATACAACAGT	AAAGACAGGA	AGTCTCAGCC	240
CATGGGACTC	TGGAGACAGG	TGGAAGGAAT	GGAGGATTGG	AAGCAGGACA	GCCAACTTCA	300
GAAGGCCAGG	GAGGACATCT	TTATGGAGAC	CCTGAAAGAC	ATCGTGGAGT	ATTACAACGA	360
CAGTAACGGG	TCTCACGTAT	TGCAGGGAAG	GTTTGGTTGT	GAGATCGAGA	ATAACAGAAG	420
CAGCGGAGCA	TTCTGGAAAT	ATTACTATGA	TGGAAGGAC	AAACTCGAG		469

## (2) INFORMATION FOR SEQ ID NO:484:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 516 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:484:

GAATTCGGCC	AAAGAGGCCT	ACTACTTCTG	TAGTCTCATC	TTGAGTAAAA	GAGAACCAG	60
CCAACTATGA	AGTTCCTTGT	CTTTGCCTTC	ATCTTGGCTC	TCATGGTTTC	CATGATTGGA	120
GCTGATTCAT	CTGAAGAGTA	TGGGTATGGC	CCTTATCAGC	CAGTCCAGA	ACAACCACTA	180
TACCCACAAC	CATACCAACC	ACAATACCAA	CCTGCCTCAA	GGTCCTCCAC	CTCCTCCAGG	240
AAAGCCACAA	GGACCACCCC	CACAAGGAGG	CAACAAACCT	CAAGGTCCCC	CACCTCCAGG	300
AAAGCCACAA	CGACCACCCC	CACAAGGAGG	CAGCAAGTCC	CGAAGTTCTC	GATCTCCTCC	360
AGGAAAGCCA	CAAGGACCAC	CCCCACAAGG	AGGCAACAAA	CCTCAAGGTC	CCCCACCTCC	420
AGGAAAGCCA	CAAGGACCAC	CCCCACAAGG	AGGCAGCAAG	TCCCGAAGTG	CCCGATCTCC	480
TCCAGGAAAG	CCACAAGGAC	CATCCACAAA	CTCGAG			516

## (2) INFORMATION FOR SEQ ID NO:485:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 357 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:485:

GAATTCGGCC	AAAGAGGCCT	ACTTCACTTC	AGCTTCACTG	ACTTCTTGAC	TCTCCTCTTG	60
AGTAAAAGGA	CTCAGCCAAC	TATGAAGTTT	TTTGTCTTTG	CTTTAGTCTT	GGCTCTCATG	120
ATTTCATGA	TTAGCGCTGA	TTACATGAA	AAGAGACATC	ATGGGTATAG	AAGAAAATTC	180
CATGAAAAGC	ATCATTCA	TCGAGAATTT	CCATTTTATG	GGGACTGTGG	ATCAAATTAT	240
CTATATGACA	ATTGATATCC	TTAGTAATCA	TGGGGCATGA	TTATAGAGGT	TTGACTGGCA	300
AATTCACTTT	TACTCATTTA	TTCTCATTCA	TCACACCGCA	AGTCTAGGCC	TCTCGAG	357

## (2) INFORMATION FOR SEQ ID NO:486:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 643 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:486: